Statistics
Canada

Statistique
Canada

MORTALITY RISK DISTRIBUTIONS: A LIFE TABLE ANALYSIS

by

Geoff Rowe

No. 19

Statistics Canada

Analytical Studies Branch

Research Paper Series

Canada

MORTALITY RISK DISTRITUTIONS:
A LIFE TABLE ANALYSIS

by

Geoff Rowe

No. 19

Social and Economic Studies Division
Statistics Canada
Fall 1988

MORTALITY RISK DISTRIBUTIONS:

A Life Table Analysis

Geoff Rowe
Social and Economic Studies Division
Statistics Canada

ABSTRACT

This study deals with the distribution of mortality risks in Canada in the period 1975-77. The description is based on nonparametric estimates of probability densities of risks of death in five year age groups. These estimates are derived from small area, age specific mortality data after adjustment for sampling variability.

Most variability in the mortality data represented variability in risks rather than instability in estimation from small populations. The variability in risks is represented in quantile life tables. These are collections of life tables, each of which is based on risk estimates at a fixed quantile of age specific risk distributions.

Summary measures of the distribution of risks are provided by estimates of quantiles of expectation of life at birth and of median age at death. Interquartile ranges for life expectancy are estimated to be 3.2 years for males and 2.8 years for females.

1. Introduction

This study is concerned with the statistical description of variability among small area, age specific probabilities of death $_nq(x)$ in the age interval $[x, x+n)$, conditional on survival to age x. In most cases, n equals five. The data, which are averages of annual data over the 1975-77 time period, comprise estimates $_n\hat{q}(x)$ for both sexes in each of 259 geographic areas in Canada. The geographic units are census divisions (CDs) which correspond to counties or other subprovincial administrative units. The exact definition differs among provinces.

The 259 CDs represent all regions in Canada with the exception of the Yukon and Northwest Territories and the most northerly part of Quebec (Nouveau-Quebec). Although the populations of many of the CDs are large in statisticians' terms, they are generally smaller than demographers usually consider adequate for construction of life tables. Demographers often assume that populations of 100,000 or less provide unstable estimates of mortality probabilities.

The study has two features which are not common in life table analysis. First, estimates of $_nq(x)$ are assumed to be subject to variability which is analogous to sampling variability. Second, the observations are assumed to come from a population which has a heterogeneous distribution of mortality risks. Heterogeneity refers to differences in risk levels and

arises from differences among individuals (e.g., diet, life style, etc.).

Each CD is treated as a sample of the Canadian population which is selected without regard to mortality conditions. Consequently, each $_n\hat{q}(x)$ is an average taken over individuals whose mortality risks are assumed to be heterogeneous. The sampling variance of $_n\hat{q}(x)$ is determined by variance of unobserved risks within CDs and by CD population size. There are marked differences in the sizes of CD populations. According to the 1976 census, the population totals of CDs in the study ranged from 1500 to 2.1 million. Thus, a description of the distribution of risk involves simultaneous use of $_nq(x)$ estimates which are not equally reliable.

The study is directed towards more fully describing mortality conditions within national populations. Within each age interval, nonparametric estimates of probability density functions of mortality risk are calculated. These estimates provide a flexible means of assessing inequality in mortality risks (e.g., by quantile based dispersion measures).

As an outcome of density estimation, period life tables are constructed by linking the estimates of age specific risk that correspond to a given quantile. These are termed quantile life tables. This approach differs from the more conventional linkage of estimates of age specific risk corresponding to a geographically defined population. The quantile life tables are calculated for grouped ages (i.e., forming abridged quantile life

tables). Abridged quantile life tables serve to illustrate an improved description of mortality for a heterogeneous national population. For purposes other than illustration, quantile life tables by single years of age would be preferred.

The emphasis in this article is on quantile life tables. A justification for basing them on CD data requires that allowance be made for differences in reliability among $_n\hat{q}(x)$. That is, the life tables should reflect heterogeneity in mortality, rather than the instability of small area estimates. A further requirement is that the variability among CD means be random, so that it is valid to describe the distribution of risks by a probability density. The two sections immediately following address these aspects of the estimation problem. Section 2 describes the $_nq(x)$ estimator and an approximation of its variance. In section 3, it is argued that geographic grouping of the data can provide a basis for estimating the distribution of risks in the Canadian population as a whole.

Estimation of the distribution of risks is in two stages: (1) section 4 describes adjustment of observations for sampling variability by a linear empirical Bayes procedure, and (2) smooth nonparametric estimates of the probability density of risks, by age, are obtained in section 5. Section 6 provides selected illustrative results for two male age groups. Section 7 presents results derived from selected quantile life tables. Finally, concluding remarks are presented in section 8.

2. Sampling Variability

Ascribing 'sampling' variance to estimates of mortality probabilities may be controversial, if death registration is complete and census undercoverage is negligible. In that case, mortality rates might be regarded as exact values. This view makes no allowance for nonrepeating and essentially unpredictable factors influencing a population's mortality in a given time period (e.g., extremes of weather, etc.). Brillinger (1986) provides a rigorous basis for describing 'natural variability', and applies one model to the mortality of the Canadian female population, 1926-82. A variety of alternative models are presented in the discussants' comments following Brillinger's paper.

Observed values $_n\hat{q}(x)_j$ for the j'th CD are represented following Chiang (1972) as (using actuarial notation):

$$_n\hat{q}(x)_j = {_nD(x)_j}/\ \{\ [\ _nP(x)_j + (n-_na(x))\ _nD(x)_j\ ]\ /\ n\ \}$$

$$= {_nD(x)_j}/K(x)_j, \tag{1}$$

where, $_nD(x)_j$ are average annual deaths in the interval 1975-77 to those aged x to x+n, $_nP(x)_j$ is the 1976 census population, and $_na(x)$ represents average years lived in the age interval by a person who died in the interval. The denominator in Eqn. 1 is an estimate of the annual population exposed to risk. In the remainder of the paper, $K(x)_j$ will be treated as if it were the true value.

The counts of deaths by age and CD were obtained from special tabulations provided by Health Division, Statistics

Canada. Population counts were obtained from the 1976 Census database. Note that population counts have been randomly rounded to multiples of five, as required for public use census tabulations.

For age groups [0,1) and [1,5), $_na(x)$ was derived from abridged male and female life tables for Canada, 1975-77 (Nagnur, 1986). For all other age groups, $_na(x)$ was assumed to take the value 2.5 (i.e., assuming a uniform distribution of deaths). The latter represents a first approximation. Other, more sophisticated, approximations are available (see, for example, Keyfitz and Frauenthal (1975)), but the appropriateness of these methods applied to small populations that are subject to substantial migration can not be readily evaluated. Consequently, only a first approximation is employed in this study, a strategy advocated by Hoem and Jensen (1982) and others.

Hereafter the subscript n, denoting the width of the age interval, is dropped and notation representing dependence on age is eliminated to simplify expressions. All subsequent analysis applies independently to each age interval.

A statistical model of the variability of $\hat{q}_j$ can be expressed in the following terms. Let Q_{ij} be a Bernoulli random variable representing possible mortality outcomes for the i'th individual in the j'th CD in the appropriate age interval. Q_{ij} takes the value 1 (representing a death) with probability q_{ij} and the value 0 with probability $1-q_{ij}$. Thus, q_{ij} represents a personal risk of death.

The expected count of deaths in the j'th CD is given by the sum (over i) of personal risks. The observed count (i.e., D_j) corresponds to the sum of realizations $\hat{q}_{ij}$ (taking value 1 if the person died and 0 otherwise) of the random variable Q_{ij}, so that Eqn. 1 provides an estimator of average risk. The true variance of the observed count (assuming independent risks) is given by:

$$Var(D_j) = \sum q_{ij} (1 - q_{ij}), \tag{2a}$$

which may be rewritten as:

$$Var(D_j) = K_j \, q._j \, (1-q._j) - \sum (q_{ij} - q._j)^2, \tag{2b}$$

where $q._j$ is the average of personal risks. Eqn. 2b shows that a binomial variance of D_j gives the maximum variance among all distributions of K_j personal risks with the average $q._j$. The accuracy of the binomial approximation may be improved by reducing the width of age intervals, as long as K_j remains large. The result depends on all q_{ij} becoming small in the limit. Note, however, that the assumption of independent risks may be too strong.

The remainder of this paper will concern itself with subsequent stages of analysis, under the assumption that the binomial approximation is adequate. Then, the approximation to the sampling variance of $\hat{q}_j$ is $\hat{q}_j(1-\hat{q}_j)/K_j$. Note that differences among K_j (i.e., sample size) contribute to differences in reliability among $\hat{q}_j$, and that is the basis for describing the variance component as 'sampling variance'.

3. Interpreting CD Mortality Differentials

The construction of regional life tables and the presentation of mortality data in atlases can give the impression that it is appropriate to treat geographically distinct populations as if they were closed to migration or as if mortality differentials somehow depend directly on region. But, since the geographic boundaries of CDs are determined without reference to mortality conditions, differences among $\hat{q}_j$ do not represent systematic differences among CD populations.

Even where some mortality risk factors are clearly associated with geography, it does not necessarily follow that geography provides an important basis for stratifying the population at risk (i.e., forming subpopulations with relatively homogeneous risks). To the extent that the risk factors involve personal habits or life style, they can be expected to vary substantially within CD populations.

One illustration of the issue is the case of malignant cutaneous melanoma. Cumulative exposure to the sun is a risk factor in some skin cancers, and purely geographic factors such as altitude and climate may be directly relevant. However, overriding influences in the case of melanoma involve individual sensitivity and intermittency of exposure. Consequently, homogeneous risk strata could be more accurately represented in terms of occupation (indoor workers at greater risk than outdoor workers), sensitivity (fair skin, tendency to burn, etc.), and

socioeconomic status (recreational tanning, winter vacations, etc.), than solely in terms of geography (Elwood et al. (1985)).

The characteristics of individuals, whose membership in one CD population rather than another is the outcome of arbitrary grouping, are a source of non-systematic variability among estimates of average risk. The grouping may correspond to a random partition of the Canadian population. Then, the distribution of CD averages will reflect the distribution of risks of typical individuals in the national population. That is, a probability density is an appropriate description of the distribution of CD averages, since the source of variability has a frequency interpretation.

4. <u>Linear Empirical Bayes Adjustment for Sampling Variability</u>

Linear empirical Bayes procedures (LEB) may be used to estimate vectors of random means. A simple form of LEB procedure can be represented as estimation of a random parameter vector $\emptyset$ by regression on a vector of observed data y (Robbins, 1983). The goal is to minimize $E\{ \emptyset(y)-\emptyset \}^2$, where $\emptyset(y)$ is a linear function of y estimating $\emptyset$. In the present application, LEB procedures involve shrinking each $\hat{q}_j$ toward the centre of the distribution. The magnitude of shrinkage is large, if the sampling variance of a given $\hat{q}_j$ is large (i.e., providing an adjustment for sampling variability).

Shrinking estimates in the direction of their overall average provides a reasonable compromise between using the overall average to represent every CD (i.e., effectively assuming

that variance between CDs is due to sampling variability) or accepting each CD estimate without modification (i.e., effectively assuming that CDs represent unrelated populations).

A variance stabilizing transformation is employed so that each $\hat{q}_j$ will have (asymptotic) sampling variance independent of $q_{.j}$. Variance stabilization is given by Anscombe's (1949) modification of the arcsin transformation:

$$\hat{t}_j = 2 \arcsin(\sqrt{(D_j + 3/8)/(K_j + 3/4)}).\tag{3}$$

Anscombe's modification gives added stability for CDs with small K, but its effect is negligible for most CDs in this application. Similar use of an arcsin transformation applied to binomial data in an LEB context is found in Efron and Morris (1975).

Variability among CD averages is represented in the model:

$$\hat{t}_j = t. + (t_j - t.) + u_j = t. + e_j,\tag{4}$$

where, $\quad t_j = 2 \arcsin(\sqrt{q_{.j}})$.

Note, $\text{Var}(e_j) = v + 1/K_j$, where $1/K_j$ represents binomial (sampling) variance of u_j given the arcsin transformation and v represents variance in excess of sampling variance (i.e., variance among CD deviates $(t_j - t.)$). Further note that $\text{Cov}(\hat{t}_j, t_j) = v$.

Estimates of $t.$ and v are obtained iteratively. The estimate $\hat{t}.$ employs the inverse of the estimated variance of each $\hat{t}_j$ as a weight:

$$\hat{t}. = \sum (\hat{t}_j \, wm_j) / \sum wm_j,\tag{5}$$

where $wm_j = 1/(\hat{v} + 1/K_j)$ with summation over m CDs. Note that if $\hat{v}$ were large enough the weights would be nearly equal (i.e., approximately independent of K_j).

An estimate of v may take the same form as would a maximum likelihood estimate assuming normality with correction factor $m/(m-1)$. A convenient computational form is:

$$\hat{v}_{i+1} = [m/(m-1)] \sum [(\hat{t}_j-\hat{t})\ wv_j]^2 / \sum wv_j, \qquad (6)$$

where $wv_j = [1/(1 + 1/(\hat{v}_i K_j))]$, with $\hat{v}_{i+1}$ a current estimate and $\hat{v}_i$ the previous estimate. Note, $E(\hat{v}) = v$ whether the deviates $(\hat{t}_j-\hat{t})$ are normal or otherwise. Given a starting value for $\hat{t}$., Eqn. 6 may be solved iteratively. Then $\hat{t}$ is estimated from Eqn. 5 with new weights, and the cycle is repeated to convergence.

Table 1 presents selected results of the variance computations. The table shows coefficients of variation (CV) and percentages of variance in excess of sampling variance (i.e., termed % Excess) for each age group. These are summary measures which provide a comparison of standard non-sampling deviations to the average and excess variance relative to sampling variance, respectively. CVs range between 8% and 28%, indicating that deviations from the average are typically substantial relative to the average. With the exception of ages 0-1, the % Excess values are all greater than 70% (i.e., indicating that most of the variability exhibited in the data is non-sampling variability). These results clearly indicate that there are important differences among the mortality levels exhibited in the CD data.

Correspondingly, the overall average is not a uniformly adequate indicator of risks experienced in the population.

Standard LEB estimates of risk (q') may be obtained from:

$$t'_j = 2 \arcsin(\sqrt{q'_j}) = \hat{t} + [\hat{v} / (\hat{v} + 1/K_j)] \, e_j, \qquad (7)$$

where the shrinkage factor applied to e_j estimates the ratio of $Cov(\hat{t}_j, t_j)$ to $Var(\hat{t}_j)$ (i.e., analogous to a regression slope). Extending the regression analogy, residual variance $Var(t'_j - t_j)$ may be expressed as $Var(t_j) - Cov(\hat{t}_j, t_j)^2/Var(t_j)$ and is estimated by $[\hat{v} / (\hat{v} + 1/K_j)] \, (1/K_j)$.

A number of authors (e.g., Louis (1984) and Gaver and O'Muircheartaigh (1987)) have expressed concern that standard LEB estimators shrink too far. For example, the observation that the sample variance $\sum (t'_j - \hat{t})^2/m-1$ is less than $\hat{v}$ implies that the empirical distribution function (EDF) of t'_j will not approximate the EDF of t_j. Louis has proposed a restricted LEB estimator which provides an approximation of the EDF of t_j that minimizes distance from the standard LEB estimates (i.e., minimizing $\sum wc_j(t^*_j - t'_j)^2$):

$$t^*_j = 2 \arcsin(\sqrt{q^*_j}) = [wc_j/(wc_j+d_1)] \, t'_j + [d_2/(wc_j+d_1)], \qquad (8)$$

where wc_j is an estimate of $1/Var(t'_j - t_j)$.

The parameters d_1 and d_2 are lagrangian multipliers chosen so that the sample average $t^*. = \sum t^*_j/m$ and the sample variance $\sum (t^*_j - t^*.)^2/m-1$ match $\hat{t}$ and $\hat{v}$, respectively. Values of d_1 for each sex and age group were obtained by a secant algorithm. At convergence approximate values of d_1 gave sample

standard deviations that differed from $\hat{v}$ by less than 0.00001 (i.e., less than 0.01% error in all cases). Parameter d_2 given d_1 was calculated by rearrangement of Eqn. 8 and provided sample averages that agreed with $\hat{t}$ exactly in each case.

The restricted LEB estimates t^*_j are reasonable estimates of risk quantiles, in as much as they are each close to the corresponding standard LEB estimate and agree closely with the estimated mean and variance of $\hat{t}_j$. Thus, the t^*_j provide a basis for approximating probability density functions of mortality risk given adjustment for differences in the reliability of $\hat{q}_j$.

5. Risk Density Function Estimates

Non-parametric estimation of a probability density function of mortality risk is intended to provide an indication of the characteristics of the distributions that would have to be accounted for in an appropriate parametric model. In that sense, this stage of analysis is exploratory. However, density estimation can also provide pseudo-likelihood location and scale estimates of mortality risk, and aid interpolation to equally spaced risk quantiles.

The goodness of fit of a simple parametric reference model of t* may be assessed from the linearity or lack of linearity in the association between t^* order statistics and expected reference quantiles. In the present application, the reference quantiles (denoted hereafter as z) are from a beta density function. The beta density has the form:

$$f\{z\} = [\ z^{b-1}\ (1-z)^{c-1}\ ]\ /\ Beta(b,c),\qquad\qquad (9)$$

where Beta(b,c) is the beta function. Beta densities can accommodate a wide variety of distributional shapes and are suitable for modeling bounded random variables (t^* is bounded by zero and Υ). The parameters b and c are chosen to match $\hat{t}/\Upsilon$ and $\hat{v}/\Upsilon^2$.

Silverman (1985) describes a smoothing cubic spline algorithm for nonparametric regression which will be used here to provide both smooth estimates of t^* and smooth estimates of first derivatives dt^*/dz. The algorithm minimizes the vector function:

$$(1-h)\ [t^*-t^*(z)]^T\ W\ [t^*-t^*(z)]\ +\ h \int t^{*\prime\prime}(z)^2\ dt^*(z),\qquad (10)$$

where t^* is a vector of order statistics, $t^*(z)$ is a vector of fitted values (from regression on z), $t^{*\prime\prime}(z)$ represents second derivatives of the fitted values, and W is a diagonal matrix of weights. The derivatives $dt^*(z)/dz$ estimate the ratio of the (known) probability of z to the probability of $t^*(z)$. Thus, non-linearities in the association between t^* and z can be represented by an empirical density differing from the fitted beta density.

The parameter h in Eqn. 10 is a smoothing parameter that differentially weights local accuracy and global smoothness. As h approaches 1.0, the algorithm represents the relation between t^* and z as a location-scale transformation (i.e., linearity implies $t^{*\prime\prime}(z)=0$). As h approaches 0.0, the algorithm implicitly includes as many parameters as there are observations (i.e., providing an interpolation where $t^*-t^*(z)=0$). An

appropriate value of h can be obtained by minimizing cross-validation errors (Silverman, 1984).

As the fitted values are estimated quantiles, W is available from the inverse of estimates of $t^*(z)$ variances:

$$Var\{t^*(z)\} = s^2 \ [\ (p(1-p)) \ / \ (m \ f\{t^*(z)\}^2) \ ], \qquad (11)$$

where p is an appropriate EDF estimate, m is the number of observations, and s^2 is a scalar (Parzen, 1979). The probability ordinates $f\{t^*(z)\}$ are obtained from the probability ordinates of z divided by the corresponding estimate of $dt^*(z)/dz$. Thus, the spline smoothing algorithm can be iterated, updating W each time from improved estimates of $f\{t^*(z)\}$ and s^2. Minimization of cross-validation errors and iterative use of Eqn. 11 led to values of h (for each age group) that were in the range 0.76 to 0.99.

Details of similar use of order statistics and model quantiles for nonparametric density estimation are found in Bofinger (1975) and Parzen (1979). The use of spline smoothing is justified by the interpretation of the spline as a moving average (kernel) smoother with a local bandwidth proportionate to the local variance (Silverman, 1984).

The estimates $t^*(z)$ are subsequently employed to construct life tables. For these purposes, $\bar{q}^*$ is obtained by inverting the arcsin transformation (i.e., $\bar{q}^*=[\sin(\ t^*(z)/2 \ )]^2$) and estimates of the probability density function of $\bar{q}^*$ are obtained by applying the Jacobian of the inverse transform to $f(t^*(z))$.

6. Illustrative Results

In this section, intermediate results are presented which illustrate the analysis. Figure 1 provides a plot of $\hat{q}$ versus $\bar{q}^*$ for males in age groups 0-1 and 50-54. These age groups were chosen because the distributions were well separated, but not so much as to completely obscure detail (i.e., by reducing the resolution of the plot). Corresponding plots of empirical density ratios are presented in Figure 2. The density ratios, estimated from the smooth first derivatives obtained above, emphasize differences that may exist between empirical mortality risk density functions and transformations of the fitted beta density functions. In that sense, the density ratios aid an evaluation of goodness of fit. Note that the plotted density ratios have been adjusted for the constant scale factor $\hat{\gamma}$.

A general index of the effects of shrinking and smoothing can be provided by the slope of a resistant line fitting $\hat{q}_j$ as a function of $\bar{q}^*_j$. The slope for males aged 0-1 is 1.33 which indicates that the dispersion of unadjusted mortality risks is typically 33% greater than the adjusted. The corresponding slope for males aged 50-54 is also 1.08. The maximum slope for males is 1.33 (ages 0-1), while the maximum for females is 1.21 (ages 35-39). The corresponding minimum values are 1.03 for males aged 20-24, and 1.06 for females aged 65-69. The effect of the adjustments on the dispersion of mortality risk estimates is not large, an observation which is not consistent

with the assumption that small area mortality estimates are generally unstable.

Figure 1 illustrates the effects of shrinking and smoothing on the $\hat{q}_j$. In age group 0-1, the upper tail (including seven or more possible outliers) has been drawn in substantially. Note that this age group has the smallest proportion of non-sampling variance of all age groups, as indicated by the % Excess values in Table 1. Consequently, this age group is more strongly influenced by shrinking than other age groups, as indicated by the resistent line slope above.

Figure 2 provides an indication of the qualitative differences there might be between the true distribution of mortality risks and the beta density that was chosen as a reference distribution. For example, density ratios for males aged 50-54 at generally close to 1.0. Thus, for this age group, the beta density may provide an adequate probability model. Perhaps more to the point, a two parameter (i.e., mean and variance) model appears adequate to describe the distribution of mortality risks for this age group. By contrast, density ratios for ages 0-1 suggest mortality risks among infants that are less skewed and have shorter tales than would be represented in a beta density. This is implied by the shape of the density ratio function, that is by the peak in the centre (values > 1.0) and low values (values < 1.0) in the tails. Density ratios similar to males aged 0-1 (i.e., indicating a lack of fit) are more common among other age groups than are density ratios similar to males

age 50-54. This would imply that three or four parameters may be required to approximate the distribution of mortality risks for most age groups.

The results presented here are sensitive to the shrinkage parameters in Eqn. 8 (i.e., d_1 and d_2) and to the choice of the smoothing parameter in Eqn. 10 (i.e., h). Thus, while the results may suggest that a four parameter model would be more appropriate tha a two parameter model, they do not provide a basis for significance tests (e.g., likelihood ratio tests).

7. Quantile Life Tables

Period life tables represent mortality conditions in terms of completed synthetic life times (e.g., forming survival curves from the product integral of age specific survival probabilities $(1-\hat{q})$). Quantile life tables are a generalization of period life tables constructed on sections through the smoothed age specific mortality distributions of $\bar{q}^*$. These sections link corresponding quantiles of different age groups. The resulting survival curves can be used to assess variability in duration of life (i.e., life expectancy at birth, or the median age at death).

Table 2 provides a comparison of previously published mortality probabilities (Nagnur, 1986 - termed AGGREGATE estimates in Table 2) with estimates employing the smoothed CD mortality probabilities and the empirical density functions. This comparison is intended to validate the life table results

presented here by demonstrating that the estimates are comparable in magnitude to more conventional estimates. Moreover, for most age groups, the order of MEAN, MEDIAN, and MODAL values clearly indicates positive skewness in the risk distributions.

However, none of the estimates in Table 2 are strictly comparable as estimates of the same measure of location. In particular, the AGGREGATE and MEAN estimates differ in as much as the former may be represented as a weighted average of $\hat{q}_j$ with K_j as weights and the latter is related to a weighted average with the reciprocals of empirical variances as weights. A weighting scheme similar to that employed in Eqn. 5 could result in markedly smaller differentials among weights compared with differentials among population sizes. Thus, the observation that estimates employing CD data are frequently higher than the more conventional AGGREGATE estimates might imply a bias in the AGGREGATE estimates. Because of the differences in these particular estimation procedures, the issue of which estimator may have the greater bias is too complex to resolve here. Nevertheless, weighting data by population size may produce inappropriate results.

Table 3 provides estimates of average and median duration of life from selected quantile life tables. These suggest that the variability in risks experienced within the Canadian population is consistent with a wide range of durations of life. Partial verification of these results is provided by comparing the low tail estimates to a group within the Canadian

population that is known to have had high mortality in the same time period. The expectation of life in the Registered Indian population in 1976 (representing about 1% of the Canadian population) was 60 years for males and 66 years for females (Rowe and Norris, 1985). Estimates based on quantile life tables are less extreme than the estimates for the Registered Indian population, which may indicate that the quantile estimates are conservative.

Table 3 provides a perspective on recent Canadian mortality trends. Over the past two decades, life expectancy in Canada has been increasing by at least one year over each five year time interval. Such increases are not large in comparison to the quantile ranges in Table 3. As such, it is worth asking whether these trends result from general improvements in health or from improvements that are specific to a part of the population. That is, would risk distributions from earlier time periods be shifted upwards or would their upper tails be stretched relative to the 1976 distributions? The distinction has bearing on our expectations for future trends in mortality.

Quantile life tables serve to focus attention on two elements of the description of mortality conditions in heterogeneous populations. That is, on the need to allow for differences among individuals of the same age, and for differences that might exist in the same group of individuals at different ages. Survival curves may be constructed by specifying both a distribution of risk at each age and specifying a path

through the succession of risk distributions. Presumably, smokers who stop smoking follow a different path than smokers who do not. Quantile life tables represent paths through the risk distributions that are in a sense determined at birth.

8. Discussion

The key findings of this study might be summarized as follows:

(a) Most of the variability among small area, age specific mortality probabilities reflects variability in mortality risks, rather than instability in small area estimates.

(b) Differences among CD averages are large enough to be of importance for descriptive or policy purposes, even after adjustment for sampling variance and after smoothing. For example, recent national mortality trends represented by changes in life expectancy might reflect contraction in the tails of the risk distributions, rather than shifts at the centre.

(c) Extreme quantile life tables provide a direct description of the range of mortality risk in a heterogeneous population by representing the best and worst synthetic life times that are consistent with the data. Quantile life tables also provide estimates of dispersion near the centre of mortality risk distributions.

The interpretation of these results depends on the assumption that CD populations could be representative samples of individuals with independent risks. This view is very different from one which explicitly or implicitly assumes region specific

effects. However, it is also a view which more nearly accords with mortality conditions in which contagious disease plays a relatively minor role while life style plays a major one.

These results might provide encouragement for further analysis of finely partitioned mortality data. Extensions might make use of partioning variables that can not be viewed as non-systematic (e.g., marital status), in addition to use of finer geographic resolution. But, for the present, replication of this work over a number of time periods might be the most promising extension.

ACKNOWLEDGEMENTS

References

Anscombe, F.J. (1948), "The transformation of poisson, binomial, and negative binomial data" Biometrika 35, 246-54.

Bofinger,E. (1975), "Estimation of a density function using order statistics" Austral.J. Statistics 17, 1-7.

Brillinger, David R. (1986), "The Natural Variability of Vital Rates and Associated Statistics" Biometrics 42, 693-734.

Chiang, C.L. (1972), "On constructing current life tables" JASA 67, 538-541.

Efron, Bradley and Morris, Carl (1975), "Data Analysis Using Stein's Estimator and Its Generalizations" JASA 70, 311-319.

Elwood,J.M., Gallagher,R.P., Hill,G.B., and Pearson, J.C.G., (1985), "Cutaneous Melanoma in Relation to Intermittent and Constant Sun Exposure - The Western Canada Melanoma Study" Int.J. Cancer 35, 427-433.

Gaver,Donald P., and O'Muircheartaigh, I.G. (1987), "Robust Empirical Bayes Analyses of Event Rates" Technometrics 29, 1-15.

Hoem, Jan M. and Jensen, Ulla Funck (1982), "Multistate Life Table Methodology: A Probabilist Critique". from Multidimensional Mathematical Demography, Kenneth C. Land and Andrei Rogers editors, Academic Press, New York.

Keyfitz, Nathan and Frauenthal, James (1975), "An Improved Life Table Method" Biometrics 31, 889-899.

Louis, Thomas A. (1984), "Estimating a Population of Parameter Values Using Bayes and Empirical Bayes Methods" JASA 79, 393-398.

Nagnur, Dhruva (1986), Longevity And Historical Life Tables: 1921-81 (Abridged) Canada and the Provinces, Catalogue 89-506, Statistics Canada, Ottawa.

Parzen, Emanuel, (1979), "Nonparametric Statistical Data Modeling" JASA 74, 105-131.

Robbins, Herbert, (1983), "Some Thoughts on Empirical Bayes Estimation" Annals of Statistics 11, 713-723.

Rowe, Geoff and Norris, M.J., (1985), "Mortality Projections of Registered Indians, 1982 to 1996" project report for Research Branch, Corporate Policy, Indian and Northern Affairs Canada, Ottawa.

Silverman, B.W. (1984), "Spline Smoothing: The Equivalent Variable Kernel Method" Annals of Statistics 12, 898-916.

Silverman, B.W. (1985), "Some Aspects of the Spline Smoothing Approach to Non-parametric Regression Curve Fitting" J.R. Statis. Soc. B, 47, 1-52.

TABLE 1

Summary Measures of Variability
in Census Division Mortality Risk
(Arcsin Transforms t_j of q_j)

AGE GROUP	MALES Coefficient of Variation	% Excess	FEMALES Coefficient of Variation	% Excess
0-1	14.7	50.1	16.3	50.1
1-4	28.4	82.1	26.2	74.6
5-9	24.8	77.8	27.5	76.1
10-14	24.7	79.7	27.3	76.0
15-19	23.0	91.3	25.0	81.5
20-24	25.4	92.8	27.3	81.0
25-29	21.4	86.7	26.0	78.7
30-34	20.5	84.0	24.0	76.9
35-39	19.1	82.7	20.2	73.6
40-44	16.4	82.7	16.8	72.3
45-49	13.7	83.5	17.6	80.7
50-54	11.5	84.2	14.3	78.5
55-59	11.3	87.6	13.2	81.9
60-64	9.6	87.8	12.1	82.7
65-69	8.7	86.7	12.5	84.9
70-74	9.3	89.5	9.7	80.6
75-79	9.0	88.2	10.1	75.7
80-84	8.2	84.6	11.1	88.0
85-89	9.2	85.0	9.1	77.7

Coefficient of variation $= 100 \sqrt{\hat{v}} / \hat{t}$

% Excess - Variance in Excess of Sampling Variance as a Percent
of Total Variance $= 100 \, (m \, v) / \sum (v + 1/K_j)$

$\hat{t}$ - estimated average, $\hat{v}$ - Non-sampling variance,
K_j - population size, m - 259 Census Divisions

Table 2

Comparison among Location Estimates
for Mortality Probabilities q

AGE	AGGREGATE	MEAN	MEDIAN	MODAL
MALES				
0-1	0.0141	0.0149	0.0146	0.0143
1-4	0.0032	0.0042	0.0039	0.0034
5-9	0.0023	0.0033	0.0032	0.0029
10-14	0.0021	0.0030	0.0028	0.0025
15-19	0.0075	0.0092	0.0087	0.0077
20-24	0.0093	0.0117	0.0111	0.0099
25-29	0.0074	0.0092	0.0088	0.0081
30-34	0.0078	0.0098	0.0095	0.0089
35-39	0.0107	0.0123	0.0118	0.0110
40-44	0.0164	0.0183	0.0178	0.0169
45-49	0.0273	0.0284	0.0279	0.0270
50-54	0.0439	0.0442	0.0437	0.0428
55-59	0.0693	0.0694	0.0687	0.0675
60-64	0.1053	0.1050	0.1054	0.1068
65-69	0.1557	0.1507	0.1499	0.1486
70-74	0.2281	0.2224	0.2219	0.2224
75-79	0.3234	0.3179	0.3165	0.3174
80-84	0.4508	0.4467	0.4453	0.4459
85-89	0.5989	0.6214	0.6207	0.6131
FEMALES				
0-1	0.0114	0.0123	0.0120	0.0115
1-4	0.0025	0.0035	0.0033	0.0032
5-9	0.0015	0.0024	0.0023	0.0020
10-14	0.0013	0.0020	0.0018	0.0016
15-19	0.0026	0.0036	0.0034	0.0030
20-24	0.0027	0.0036	0.0034	0.0030
25-29	0.0028	0.0036	0.0034	0.0028
30-34	0.0038	0.0047	0.0044	0.0039
35-39	0.0059	0.0072	0.0070	0.0067
40-44	0.0091	0.0103	0.0101	0.0097
45-49	0.0152	0.0151	0.0147	0.0144
50-54	0.0218	0.0227	0.0225	0.0220
55-59	0.0339	0.0342	0.0338	0.0330
60-64	0.0512	0.0513	0.0511	0.0515
65-69	0.0800	0.0795	0.0794	0.0804
70-74	0.1267	0.1282	0.1279	0.1287
75-79	0.2010	0.2037	0.2024	0.2010
80-84	0.3222	0.3318	0.3324	0.3337
85-89	0.4866	0.4965	0.4919	0.4817

AGGREGATE - as published in Nagnur (1986) - based on the ratios
    of total deaths to total population.
MEAN, MEDIAN, MODAL - as determined by empirical Bayes estimates
    of mortality risk and empirical probability densities.

TABLE 3

Quantile Estimates of Duration of Life

Quantile	Life Expectancy at Birth		Median Age at Death	
	MALES	FEMALES	MALES	FEMALES
1%	63.4	71.4	68.9	77.0
5%	65.5	73.4	70.5	78.4
10%	66.5	74.3	71.2	79.1
25%	68.2	75.7	72.5	80.2
50%	69.8	77.1	73.8	81.1
75%	71.4	78.5	75.2	82.1
90%	72.8	79.7	76.4	83.0
95%	73.7	80.5	77.1	83.6
99%	75.3	82.0	78.4	84.8
MEAN	69.6	77.0	73.7	81.0
MODAL	70.2	77.4	74.0	81.2
AGGREGATE	70.26	77.70	73.8	81.3

MEAN, MODAL estimates obtained from mortality probabilities in
    Table 2.
AGGREGATE - life expectancy as published in Nagnur (1986)-
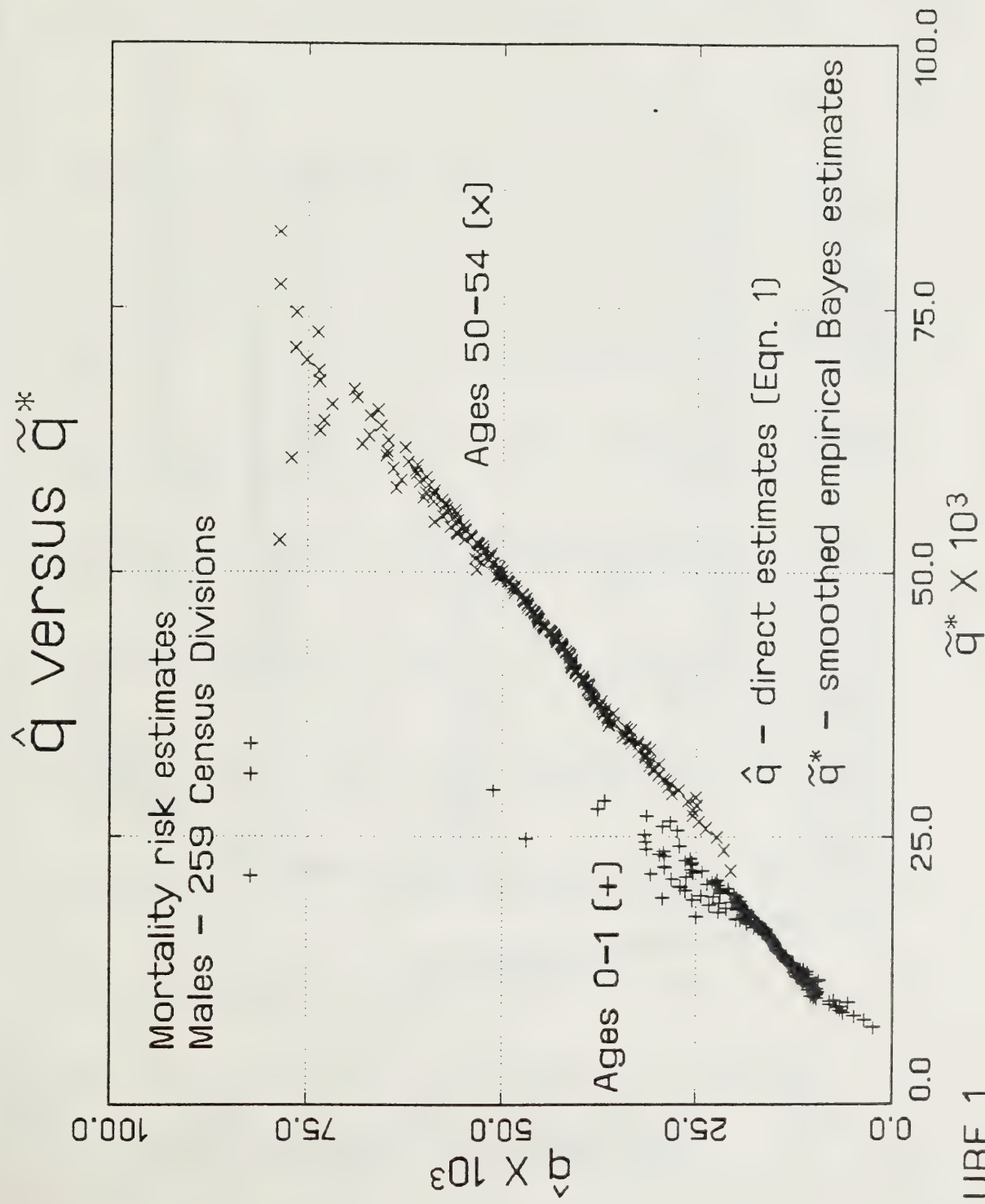   median ages at death interpolated from the published
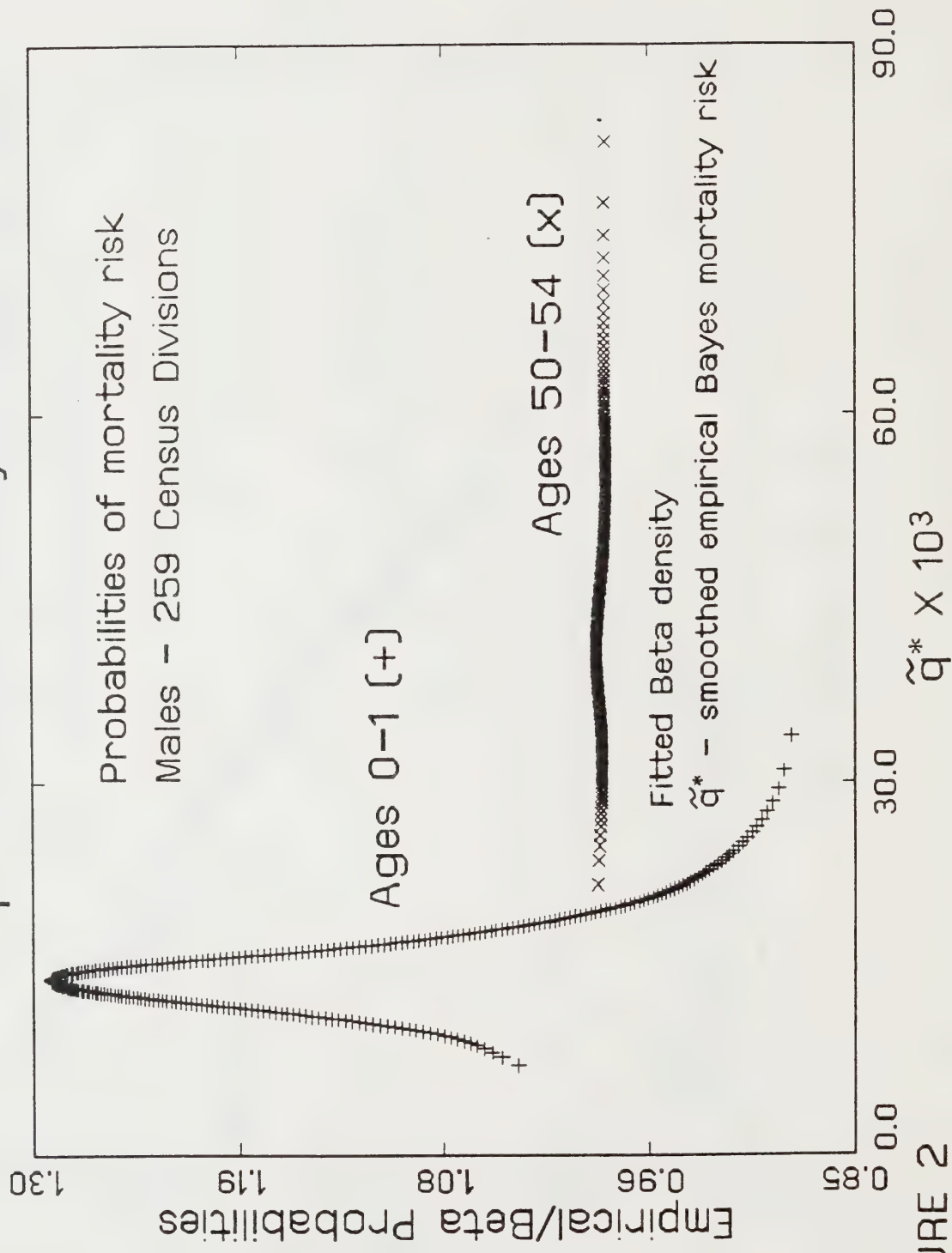   survivor curve.

q̂ versus q̃*

q̂ versus q̃*

Mortality risk estimates
Males – 259 Census Divisions

Ages 50–54 [×]

Ages 0–1 [+]

q̂ – direct estimates [Eqn. 1]
q̃* – smoothed empirical Bayes estimates

$\tilde{q}^* \times 10^3$

$\hat{q} \times 10^3$

FIGURE 1

Empirical Density Ratios

Probabilities of mortality risk
Males – 259 Census Divisions

Ages 0–1 [+]

Ages 50–54 [x]

Fitted Beta density
$\tilde{q}^*$ – smoothed empirical Bayes mortality risk

$\tilde{q}^* \times 10^3$

Empirical/Beta Probabilities

FIGURE 2

ANALYTICAL STUDIES BRANCH

RESEARCH PAPER SERIES

11. ENERGY ANALYSIS OF CANADIAN EXTERNAL TRADE: 1971 and 1976 by K.E. Hamilton

12. NET AND GROSS RATES OF LAND CONCENTRATION by Ray Bollman and Philip Ehrensaft

13. CAUSE-DELETED LIFE TABLES FOR CANADA (1921 to 1981): AN APPROACH TOWARDS ANALYSING EPIDEMIOLOGIC TRANSITION by Drhuva Nagnur and Michael Nagrodski

14. THE DISTRIBUTION OF THE FREQUENCY OF OCCURRENCE OF NUCLEOTIDE SUBSEQUENCES BASED ON THEIR OVERLAP CAPABILITY by Jane F. Gentleman and Ronald C. Mullin

15. IMMIGRATION AND THE ETHNOLINGUISTIC CHARACTER OF CANADA AND QUEBEC by Réjean Lachapelle

16. INTEGRATION OF CANADIAN FARM AND OFF-FARM MARKETS AND THE OFF-FARM WORK OF WOMEN, MEN AND CHILDREN by Ray Bollman and Pamela Smith

17. WAGES AND JOBS IN THE 1980s: CHANGING YOUTH WAGES AND THE DECLINING MIDDLE by J. Myles, G. Picot and T. Wannell

18. A PROFILE OF FARMERS WITH COMPUTERS by Ray D. Bollman

19. MORTALITY RISK DISTRIBUTIONS: A LIFE TABLE ANALYSIS by Geoff Rowe

For further information, contact the Chairperson, Publication Review Committee, Analytical Studies Branch, R.H. Coats Bldg., 24th Floor, Statistics Canada, Tunney's Pasture, Ottawa, Ontario K1A 0T6.